

# Goals and Organization for Discussion on Data Quality. DØ Beaune Workshop, 16-20 June 2003

9th May 2003

## Goals

The physics output of DØ depends strongly on the amount of luminosity that is recorded. An important factor in here is the quality of the data that we write to tape and reconstruct: bad data, which cannot reliably be used in an analysis is lost luminosity. Another factor is that the good data should be of the highest achievable quality, as free as is attainable from detector imperfections and reconstruction inadequacies.

The goal of this workshop is to review our current approach to data quality, and make plans for the future. This clearly also requires an idea of what quality is required, preferably from the point of view of a physics analysis. At the end of the meeting, a plan should be ready for attacking data quality from the lowest level in the control room, to the highest level in a physics analysis. The organization of a system to efficiently communicate and attack quality issues through these levels should have started.

## Organization

The workshop will begin with a short series of introductory plenary talks. After that, there will be three days of parallel sessions. On the last day, the result of the work done in the parallels will be summarized in a plenary presentation.

## Plenary

A first set of plenary talks will be devoted to specifications on data quality for benchmark physics analysis. This clearly requires work before the workshop. Some analyses are beginning to be developed enough, with at least partial identification of systematic uncertainties, to provide a basis for making a specification. Other analyses may not be worked on yet, but experience from Run I, Run II workshops or other experiments may help.

A second and third plenary session will be filled by presentations from object ID and detector groups. This may include a review of activities but should focus on current efforts on data quality monitoring. Again, this requires work before the workshop, e.g. by the group leaders, to prepare an overview of current activities and to think about future plans. Also,

existing procedures to solve data quality problems should be addressed. In a fourth plenary session, global (online) monitoring and offline monitoring will be reviewed.

## Parallel 1

The first series of parallel sessions will be organized along the lines of sub-detector systems. In these sessions, both online and offline topics should be discussed. Clearly, this goes beyond detector, ID or physics group boundaries since different groups use the same detector systems. Topics to consider:

- Review and define quality requirements: raw / reconstructed data, stability
- Review current data quality procedures:
  - Online: shifter, detector expert
  - Offline: reco certification, physics
  - Timescales: run, luminosity block, event
  - Quality information storage / availability / format / communication
  - Trouble shooting procedures
- Review quality problems that occurred:
  - How and when caught
  - How and when solved
- Plans for the future: how to improve data quality / communication

The following groups will meet in these parallel sessions. For each, a (non exhaustive) list of detector / ID groups is given which can and should contribute to that session. Physics groups are encouraged to report their problems and thoughts in the appropriate session.

**Track** CFT, SMT, tracking algo., vertex algo., alignment, b-ID,  $\mu$ -ID,  $\tau$ -ID

**Cal** Calorimeter, Calorimeter task force, EM-ID, JET/MET, JES

**Muon** Muon detector,  $\mu$ -ID, MET

**C/FPS** PS detector, EM-ID, JES

**Beam** Everything involving beam “quality”: halo / background rates (Lumi, FPD), (real time) beam position (trigger, tracking, vertexing), (real time) beam profile (Lumi, tracking)

**FPD** Detector

**L1/L2** Cal, Muon, CTT, STT, PS

**L3** DAQ, L3 filtering and monitoring, streaming

**Lumi** Detector, monitoring, offline access, uncertainty

For each of these subgroups, the sessionleaders are expected not only to organize the session, but also to prepare a summary of the activity in their session for the second series of parallel sessions.

## **Parallel 2**

A second series of parallel sessions is meant to bring together the different sub-detector groups, split between online and offline.

In a first session, the ideas brought up in the detector system sessions will be reviewed and discussed. This should lead to a plan for efficiently attacking data quality issues and establishing data evaluation procedures, including communication lines. A second session will be devoted to higher level on- and offline issues, such as: data quality information storage, databases, streaming, luminosity, offline shifts, versioning, re-reconstruction of data and combining data samples.

These sessions are clearly meant to be even more discussion sessions than the first series of parallels.